Portfolio Data Analytics
2nd

Iowa Liquor Sales Insight

TAKEN FROM CASE STUDY
RevoU Mini Course - Data Analytics

DHEA AMALIA LUTFIANI
(27 JUNE – 8 JULY)

Case Study Instructions

QUESTION

Table of interest :

‘bigquery-public-data.iowa_liquor_sales.sales’

1. Look at this data and start thinking. List down 3 trends/points that you want to show.
2. From here, try to explore the data and make changes, filter, and prepare the data that you need.
3. Create some visualizations or dashboard with the best type of chart you have learned.

   The easiest is with Google Data Studio or Google Sheets.
4. Then, make 1-2 slides from the Graphs with the insights you got to present your findingsto the stakeholders

   (read this article from HBR)

Preview Data

Schema from the Dataset :

Field name	Type	Mode	Collation	Policy Tags	Description
invoice_and_item_number	STRING	NULLABLE			Concatenated invoice and line number associated with the liquor order. This provides a unique identifier for the individual liquor products included in the store order.
date	DATE	NULLABLE			Date of order
store_number	STRING	NULLABLE			Unique number assigned to the store who ordered the liquor.
store_name	STRING	NULLABLE			Name of store who ordered the liquor.
address	STRING	NULLABLE			Address of store who ordered the liquor.
city	STRING	NULLABLE			City where the store who ordered the liquor is located
zip_code	STRING	NULLABLE			Zip code where the store who ordered the liquor is located
store_location	STRING	NULLABLE			Location of store who ordered the liquor. The Address, City, State and Zip Code are geocoded to provide geographic coordinates. Accuracy of geocoding is dependent on how well the address is interpreted and the completeness of the reference data used.
county_number	STRING	NULLABLE			Iowa county number for the county where store who ordered the liquor is located
county	STRING	NULLABLE			County where the store who ordered the liquor is located
category	STRING	NULLABLE			Category code associated with the liquor ordered
category_name	STRING	NULLABLE			Category of the liquor ordered.
vendor_number	STRING	NULLABLE			The vendor number of the company for the brand of liquor ordered
vendor_name	STRING	NULLABLE			The vendor name of the company for the brand of liquor ordered
item_number	STRING	NULLABLE			Item number for the individual liquor product ordered.

Preview Data

Schema from the Dataset :

item_description	STRING	NULLABLE	Description of the individual liquor product ordered.
pack	INTEGER	NULLABLE	The number of bottles in a case for the liquor ordered
bottle_volume_ml	INTEGER	NULLABLE	Volume of each liquor bottle ordered in milliliters.
state_bottle_cost	FLOAT	NULLABLE	The amount that Alcoholic Beverages Division paid for each bottle of liquor ordered
state_bottle_retail	FLOAT	NULLABLE	The amount the store paid for each bottle of liquor ordered
bottles_sold	INTEGER	NULLABLE	The number of bottles of liquor ordered by the store
sale_dollars	FLOAT	NULLABLE	Total cost of liquor order (number of bottles multiplied by the state bottle retail)
volume_sold_liters	FLOAT	NULLABLE	Total volume of liquor ordered in liters. (i.e. (Bottle Volume (ml) x Bottles Sold)/1,000)"
volume_sold_gallons	FLOAT	NULLABLE	Total volume of liquor ordered in gallons. (i.e. (Bottle Volume (ml) x Bottles Sold)/3785.411784)"

From this data preview, we can find out the description of each column and know which columns can be used to answer problems or which columns can be useful for finding new insights.

Preview Data

Preview the Data :

SCHEMA DETAILS PREVIEW

Row	invoice_and_item_number	date	store_number	store_name	address	city	zip_code	store_location	county_number	county
1	INV-31859900152	2020-11-11	2629	Hy-Vee Food Store #2 / Council Bluffs	1745 Madison Ave	Council Bluffs	51503.0	POINT (-95.825137 41.242732)	78	POTTAWATTA
2	INV-31888100019	2020-11-12	4959	Bani's	2128 College St	Cedar Falls	50613.0	POINT (-92.455801 42.518018000000005)	7	BLACK HAWK
3	INV-08543300047	2017-11-08	3447	Sam's Club 6432 / Sioux City	4201 S. York St.	Sioux City	51106	POINT (-96.37082 42.43407)	97	WOODBURY
4	S30413700008	2016-01-28	4214	New Star Liquor / Lafayet / Waterloo	1309 LAFAYETTE ST	WATERLOO	50703	POINT (-92.3272 42.494092)	7	Black Hawk
5	S30214200009	2016-01-15	5100	Sam's Food	648 N MARQUETTE ST	DAVENPORT	52802	POINT (-90.590879 41.526469)	82	Scott
6	S30501900059	2016-02-02	4167	Iowa Street Market, Inc.	1256 IOWA ST	DUBUQUE	52001	POINT (-90.668138 42.504959)	31	Dubuque
7	INV-16588500106	2018-12-27	2614	Hy-Vee #3 Food & Drugstore / Davenport	1823 E Kimberly Rd	Davenport	52807	POINT (-90.548919 41.55678100000001)	82	SCOTT
8	S22205000040	2014-11-05	3819	Quillins Decorah	915 SHORT ST	DECORAH	52101	POINT (-91.79777100000001 43.291023)	96	Winneshiek
9	INV-36386200001	2021-05-06	2647	Hy-Vee #7 / Cedar Rapids	5050 Edgewood Rd	Cedar Rapids	52411.0	POINT (-91.701581 42.030129)	57	LINN
10	S06296300001	2012-06-27	3990	Cork and Bottle / Oskaloosa	309 A AVE WEST	OSKALOOSA	52577	POINT (-92.648153 41.296228)	62	Mahaska

SCHEMA DETAILS PREVIEW

Row	category	category_name	vendor_number	vendor_name	item_number	item_description	pack	bottle_volume_ml	state_bottle_cost
1	1062300.0	Aged Dark Rum	421	SAZERAC COMPANY INC	42167	Myers's Original Dark Rum	12	1000	13.0
2	1700000.0	Temporary & Specialty Packages	260	DIAGEO AMERICAS	101217	Captain Morgan OSR PET 6/1.75l w/ 50ml CM Sliced Apple	6	1750	18.0
3	1092100.0	Imported Distilled Spirit Specialty	434	LUXCO INC	75087	Juarez Gold Dss	12	1000	5.17
4	1051100.0	APRICOT BRANDIES	115	Constellation Wine Company, Inc.	56803	Paul Masson Red Berry Grande Amber Brandy	24	200	1.96
5	1051140.0	PEACH BRANDIES	115.0	Constellation Wine Company, Inc.	56193	Paul Masson Peach Grande Amber Brandy	24	200	1.96
6	1031100.0	100 PROOF VODKA	297.0	Laird And Company	35917	Five O'clock Vodka	12	1000	4.17
7	1081100.0	Coffee Liqueurs	370	PERNOD RICARD USA	67527	Kahlua Coffee	12	1000	14.99
8	1701100.0	DECANTERS & SPECIALTY PACKAGES	322	Prestige Wine and Spirits Group	594	Opulent w/Shaker	3	1750	15.0
9	1012300	Single Malt Scotch	370	PERNOD RICARD USA	5038	Glenlivet 12YR	6	1750	44.97
10	1081700.0	DISTILLED SPIRITS SPECIALTY	434	Luxco-St Louis	75087	Juarez Gold Dss	12	1000	4.75

Preview Data

Preview the Data :

SCHEMA DETAILS PREVIEW

Row	item_number	item_description	pack	bottle_volume_ml	state_bottle_cost	state_bottle_retail	bottles_sold	sale_dollars	volume_sold_liters	volume_sold_gallons
1	42167	Myers's Original Dark Rum	12	1000	13.0	19.5	6	117.0	6.0	1.58
2	101217	Captain Morgan OSR PET 6/1.75l w/ 50ml CM Sliced Apple	6	1750	18.0	27.0	36	972.0	63.0	16.64
3	75087	Juarez Gold Dss	12	1000	5.17	7.76	60	442.8	60.0	15.85
4	56803	Paul Masson Red Berry Grande Amber Brandy	24	200	1.96	2.94	24	70.56	4.8	1.27
5	56193	Paul Masson Peach Grande Amber Brandy	24	200	1.96	2.94	24	70.56	4.8	1.27
6	35917	Five O'clock Vodka	12	1000	4.17	6.26	60	375.6	60.0	15.85
7	67527	Kahlua Coffee	12	1000	14.99	22.49	4	89.96	4.0	1.05
8	594	Opulent w/Shaker	3	1750	15.0	22.5	3	67.5	5.25	1.39
9	5038	Glenlivet 12YR	6	1750	44.97	67.46	24	1619.04	42.0	11.09
10	75087	Juarez Gold Dss	12	1000	4.75	7.13	96	684.48	96.0	25.36

Defining Question

List down trends/points that you want to show :

1. How was the monthly trend of sales?
2. How was our sales in Iowa's City?
3. Which category sells the most liquor?
4. At which day in a week most of customers buy a liquor?
5. Is liquor with large volume has a better sales compared with small volume?

Exploring Data

Include SQL with BigQuery and Data Cleaning

- Look for the data with BigQuery > SQL

1. For some columns with string fields such as in the "city, country, and others" column, the case for this data is that their filling is not the same. Some use only capital letters and some use capital letters at the beginning of the word. Therefore, it is necessary to do Data Cleaning to equalize the data entry string in that column.

- The following is a query to see one of these columns:

```
1  SELECT  DISTINCT city
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  ORDER BY city
```

- And the result :

Row	city
1	*null*
2	ACKLEY
3	ADAIR
4	ADEL
5	AFTON
6	AKRON
7	ALBIA
8	ALDEN

Row	city
29	Ackley
30	Adair
31	Adel
32	Afton
33	Akron
34	Albert City
35	Albia
36	Albion

Exploring Data

Include SQL with BigQuery and Data Cleaning

- Syntax to convert all strings to capital, for example as follows in the city column :

```
1  SELECT DISTINCT UPPER(city)
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  ORDER BY 1
```

- And the result is that all the contents or values of the city column are only CAPITALIZED and no value is the same.

2. For the category column, it has the same data or value entries but the writing is different. This is also where data cleaning is needed.

- The following is a query to view those columns :

```
1  SELECT DISTINCT category
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  ORDER BY 1
```

- And the result :
  There are 2 values that are the same but written differently, namely ".0" behind.

Row	category
1	*null*
2	1011000
3	1011100
4	1011100.0
5	1011200
6	1011200.0
7	1011250.0
8	1011300

Exploring Data

Include SQL with BigQuery and Data Cleaning

- Syntax to remove ".0" in the category column:

```
1  SELECT DISTINCT REPLACE(category, ".0", "")
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  ORDER BY 1
```

- And the result :
  No more data equal to writing ".0"

Row	f0_
1	*null*
2	1011000
3	1011100
4	1011200
5	1011250
6	1011300
7	1011400
8	1011500

3. For columns that have more than 1 value, data cleaning is also required. As in the following category_name column, with the same number but different values.

- The following is a query to view the column :

```
1  SELECT DISTINCT REPLACE(category, ".0", ""), UPPER(category_name)
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  ORDER BY 1,2
```

Exploring Data

Include SQL with BigQuery and Data Cleaning

- And the result :

Row	f0_	f1_
1	*null*	*null*
2	1011000	AMERICAN WHISKIES
3	1011100	BLENDED WHISKIES
4	1011200	STRAIGHT BOURBON WHISKIES
5	1011250	SINGLE BARREL BOURBON WHISKIES
6	1011300	SINGLE BARREL BOURBON WHISKIES
7	1011300	TENNESSEE WHISKIES
8	1011400	BOTTLED IN BOND BOURBON

- Syntax to take just one value from the same nomor identity :

```
1  SELECT DISTINCT REPLACE(category, ".0", ""), ANY_VALUE(UPPER(category_name))
2  FROM `bigquery-public-data.iowa_liquor_sales.sales`
3  GROUP BY 1
4  ORDER BY 1,2
```

- And the result :
  No 2 or more numbers with the same identity have different category_name values

Row	f0_	f1_
1	*null*	*null*
2	1011000	AMERICAN WHISKIES
3	1011100	BLENDED WHISKIES
4	1011200	STRAIGHT BOURBON WHISKIES
5	1011250	SINGLE BARREL BOURBON WHISKIES
6	1011300	SINGLE BARREL BOURBON WHISKIES
7	1011400	BOTTLED IN BOND BOURBON
8	1011500	STRAIGHT RYE WHISKIES
9	1011600	STRAIGHT RYE WHISKIES
10	1011700	CORN WHISKIES
11	1011800	IOWA DISTILLERY WHISKIES
12	1012000	IMPORTED WHISKIES
13	1012100	CANADIAN WHISKIES

Exploring Data

- And finally, the following is a combined query that results from data exploration:
- Where there is an aggregate function for grouping bottle sizes, taking only one data value for the identity number value, and cleaning data that has been described previously, such as removing the string ".0".
- In addition, there is also a separation of location data between its longitude and latitude, as well as eliminating null values in certain columns.

```sql
SELECT
  date,
  invoice_and_item_number,
  city,
  REPLACE(zip_code, ".0", "") zip_code,
  CASE WHEN bottle_volume_ml BETWEEN 0 AND 500 THEN 'a.Small Size (0-500 ml)'
    WHEN bottle_volume_ml BETWEEN 501 AND 1000 THEN 'b.Medium Size (501-1000 ml)'
    WHEN bottle_volume_ml > 1000 THEN 'c.Large Size (>1000 ml)'
    ELSE null
  END AS category_size_bottle,
  bottles_sold,
  sale_dollars,
  volume_sold_liters,
  store_number,
  county_number,
  REPLACE(category, ".0","") category,
  vendor_number,
  item_number,
  store_location,
  ST_GEOGFROMTEXT(store_location) point_location,
    ST_X(ST_GEOGFROMTEXT(store_location)) as longitude,
    ST_Y(ST_GEOGFROMTEXT(store_location)) as latitude,
  ANY_VALUE(UPPER(store_name)) store_name,
  ANY_VALUE(UPPER(county)) county,
  ANY_VALUE(UPPER(category_name)) category_name,
  ANY_VALUE(UPPER(vendor_name)) vendor_name,
  ANY_VALUE(UPPER(item_description)) item_description,
FROM `bigquery-public-data.iowa_liquor_sales.sales`
WHERE county_number is not null
AND store_number is not null
AND category is not null
AND vendor_number is not null
AND item_number is not null
AND date BETWEEN '2021-01-01' AND '2021-12-31'
GROUP BY 1,2,3,4,5,6,7,8,9,10,11,12,13,14
```

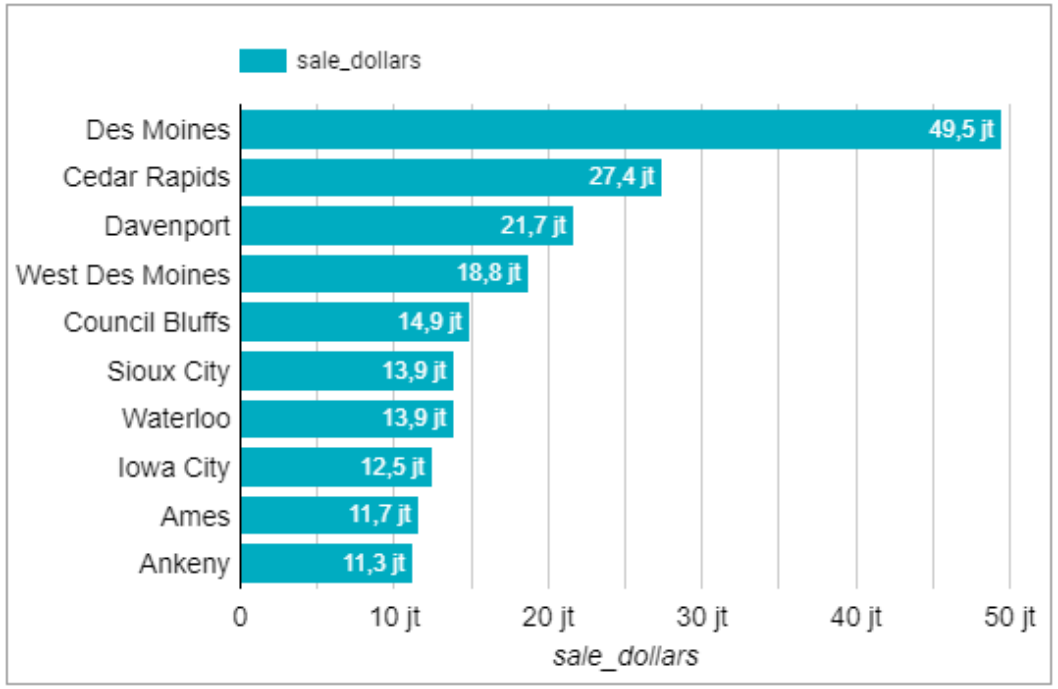
Visualization with Insight

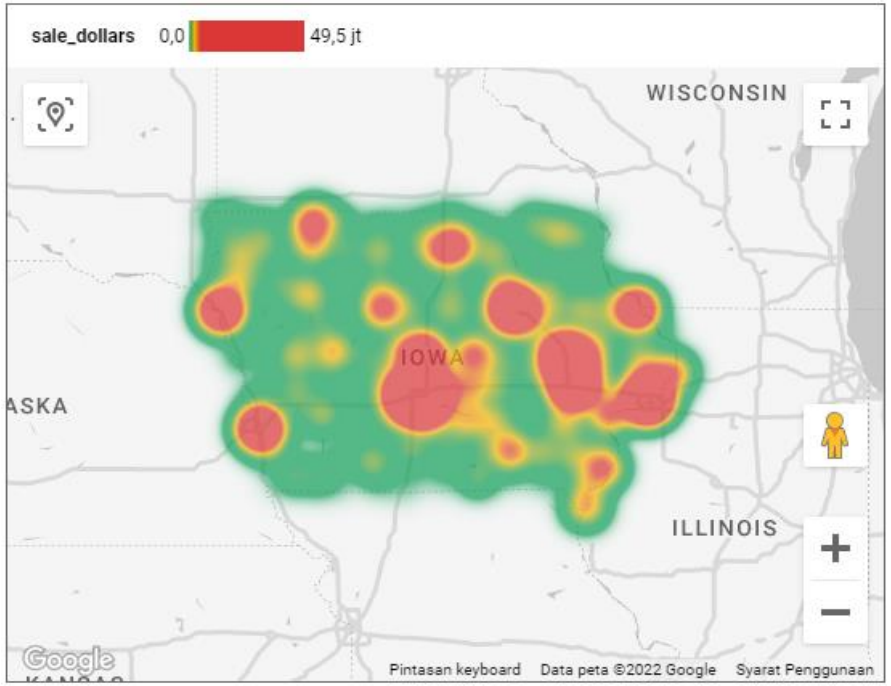
1. How was the monthly trend of sales?

- From the diagram, it can be seen that sales in 2021 tend to increase from the beginning to the end of the year. Where the highest sales are in December 2021.
- Then for the largest ratio of sales per volume occurred in October 2021 and the smallest ratio occured in January 2021.
- It means volumes with higher prices were sold better in October 2021 .

Visualization with Insight

2. How was our sales in Iowa's City?

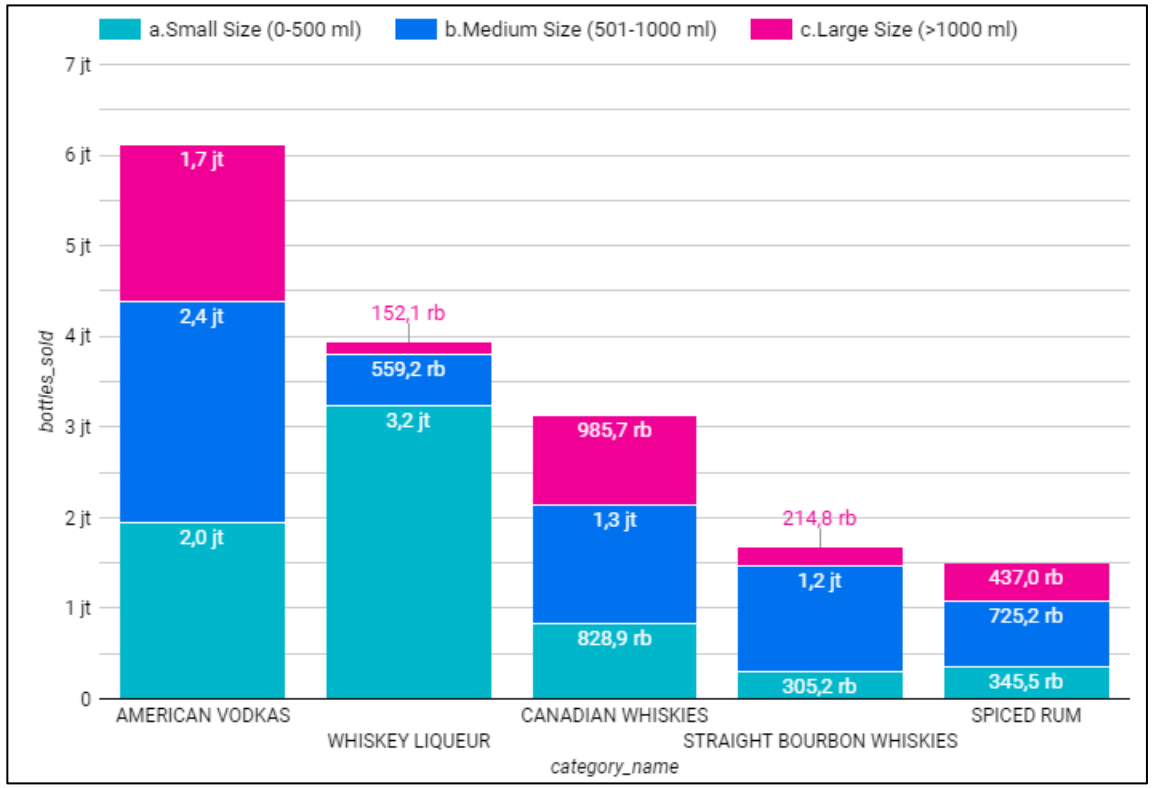




- From the bar chart above Des Moines is the city with the largest liquor sales in Iowa with nearly $50 million. And sales in Des Moines reached 1.8 times the sales of the second rank, namely Cedar Rapids.
- Then the following heatmap shows the 10 cities with the highest sales in Iowa, indicated in red and the other cities with low sales in green.

Visualization with Insight

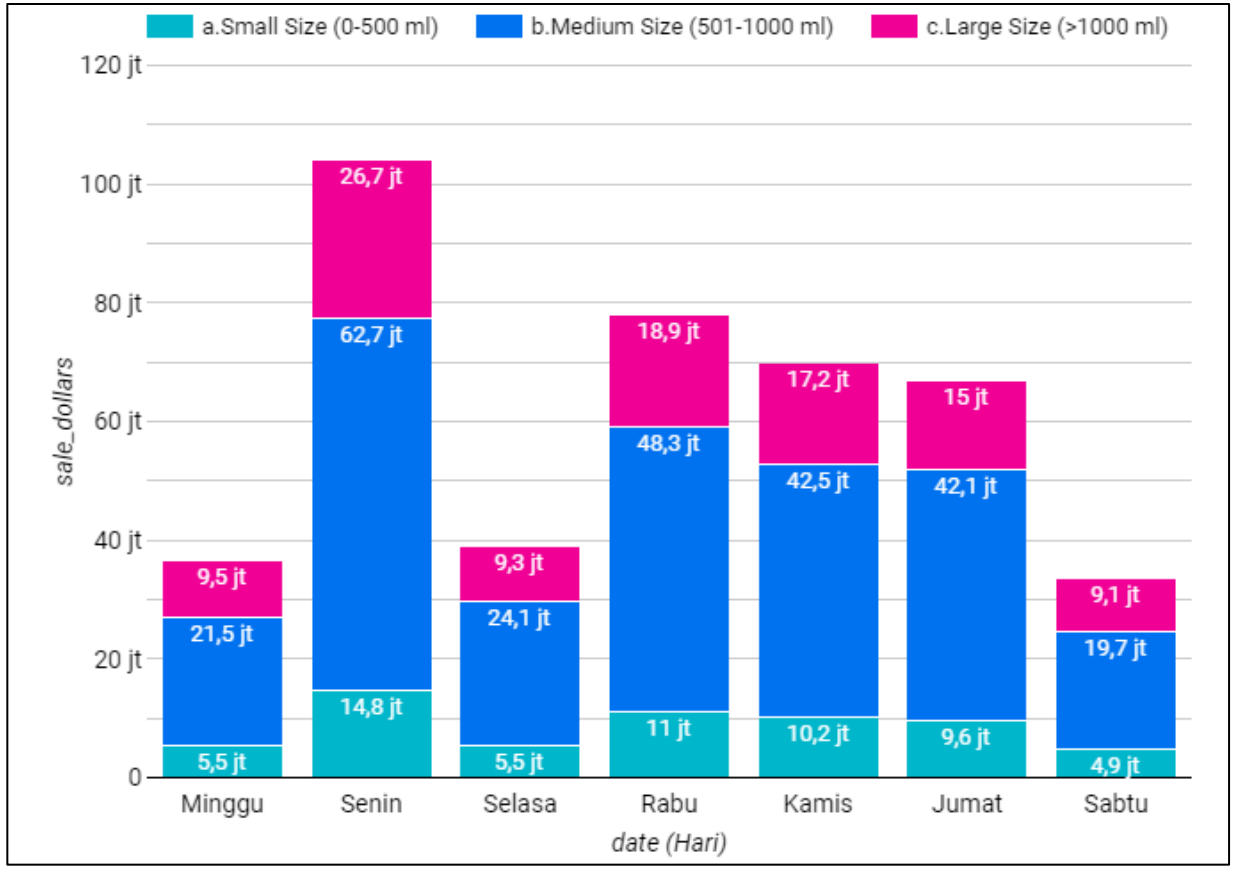
3. Which category sells the most liquor?



- The following chart shows the top 5 categories or types of liquor purchased in Iowa.
- We know that the most sold liquor is American Vodkas with a total of more than 6 million, but with almost the same number of purchases per size.
- Sales of liquor type Whiskey Liqueur small size is the most sold liquor to 3.2 million dollars.

Visualization with Insight

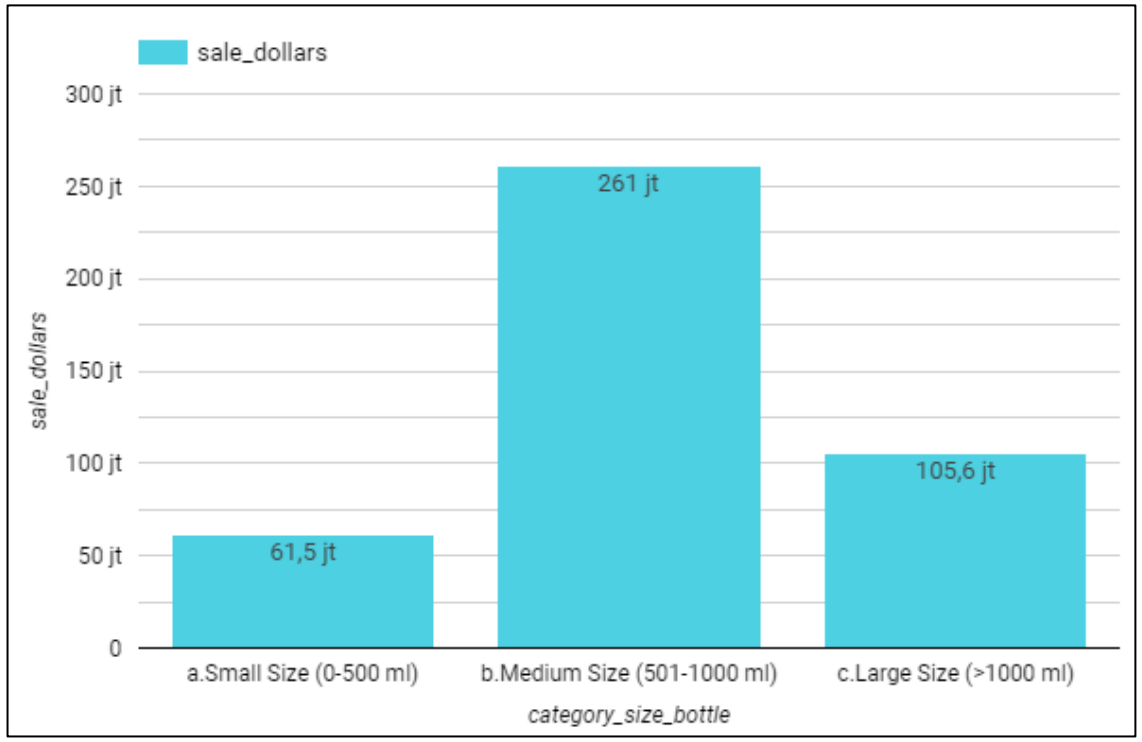
4. At which day in a week most of customers buy a liquor?



- Monday was the day with the largest amount of liquor sales reaching 26.7 million dollars and Saturday was the lowest selling day with less than 10 million dollars.
- It can be concluded that liquor sales are fluctuating every day.

Visualization with Insight

5. Is liquor with large volume has a better sales compared with small volume?



- When compared, it turns out that liquor sales for the large size are greater than the small size.
- However, the largest sales are obtained from the sale of liquor with medium size, which is 501 ml to 1000 ml.

THANK YOU

DHEA AMALIA LUTFIANI

(27 JUNE – 8 JULY)